

Final pre-publication version.

For final published version see:

McCreesh KM, Crotty JM, Lewis JS 2015 Acromiohumeral distance measurement in rotator cuff tendinopathy: is there a reliable, clinically applicable method? *Br J Sports Med*; 49:298-305 doi:10.1136/bjsports-2012-092063

**Acromiohumeral distance measurement in rotator cuff tendinopathy:
is there a reliable, clinically applicable method? A systematic review**

Authors:

Corresponding Author

Karen M McCreesh

Dept of Clinical Therapies, University of Limerick, Limerick, Ireland

Email: karen.mccreesh@ul.ie

Phone: +353 61234233 **FAX:** +353 61234251

Co Authors:

James M Crotty

Dept of Radiology, University Hospital Limerick, Limerick, Ireland

Jeremy S Lewis

**Department of Allied Health Professions and Midwifery, School of Health and
Social Work, University of Hertfordshire, Hatfield,UK**

**Musculoskeletal Services, Health at the Stowe, Central London Community
Healthcare NHS Trust London, UK**

Abstract

Background: Narrowing of the subacromial space has been noted as a common feature of rotator cuff (RC) tendinopathy, and has been implicated in the development of symptoms, and forms the basis for some surgical and rehabilitation approaches. Various radiological methods have been used to measure the subacromial space, which is represented by a two-dimensional measurement of acromio-humeral distance (AHD). A reliable method of measurement could be used to assess the impact of rehabilitation or surgical interventions for RC tendinopathy, however there are no published reviews assessing the reliability of AHD measurement.

Objectives: The aim of this review was to systematically assess the evidence for the intra- and inter-rater reliability of radiological methods of measuring AHD, in order to identify the most reliable method for use in rotator cuff tendinopathy.

Study appraisal and synthesis: An electronic literature search was carried out and studies describing the reliability of any radiological method of measuring AHD in either healthy or RC tendinopathy groups were included. Eighteen studies met the inclusion criteria, and were appraised by two reviewers using the Quality Appraisal for reliability Studies (QAREL) checklist.

Results: Eight studies were deemed to be of high methodological quality. Study weaknesses included lack of tester blinding, inadequate description of tester experience, lack of inclusion of symptomatic populations, poor reporting of statistical methods and unclear diagnosis. There was a strong evidence for the reliability of ultrasound for measuring AHD, with moderate evidence for Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) measures, and conflicting evidence for radiographic methods. Overall there was a lack of research in rotator cuff tendinopathy populations, with only 6 studies including participants with shoulder pain.

Conclusion: The results support the reliability of ultrasound and CT or MRI for the measurement of AHD, however more studies in symptomatic populations are required. The reliability of AHD measurement using radiographs has not been supported by the studies reviewed.

Keywords: acromiohumeral distance, rotator cuff, reliability

Reliability of acromiohumeral distance measurement in rotator cuff tendinopathy: a systematic review

BACKGROUND

Shoulder pain is a common musculoskeletal condition, with point prevalence rates of between 7 and 26% in adults¹. The most common source of adult shoulder pain is rotator cuff (RC) tendinopathy, which is a multi-factorial condition². The supraspinatus tendon, which runs in the subacromial space, is most commonly affected by pathological change. Narrowing of the subacromial space has been variously ascribed to: loss of RC function leading to superior migration of the humeral head³, altered acromial morphology⁴, or postural alterations⁵. This phenomenon of a reduced subacromial space, and proposed resulting impingement of the RC tendons and subacromial bursa, has been widely implicated in the development of degenerative RC pathology and pain in both athletic and non-athletic populations⁶.

The size of the subacromial space is commonly quantified by the measurement of distance between the acromion and the humeral head termed the 'acromiohumeral distance' (AHD), using a variety of different radiological methods, including radiographs, CT scans, MRI, and ultrasound. Studies of AHD in asymptomatic shoulders have reported ranges of AHD from 6 to 12 mm in the neutral position⁷⁻⁹. The source of this variation may be due to inter-individual variability, or the variety of measurement protocols used. Generally, AHD is found to be reduced as the arm moves into abduction up to 90°¹⁰ and has been shown to be influenced by muscle contraction¹¹, and by muscle fatigue¹². In symptomatic populations, radiographic

studies have suggested an AHD cut-off point of 6-7mm to indicate the presence of a significant rotator cuff tear, with recent work by Goutallier et al¹³ suggesting that a 6mm cut-off is indicative of a large tear, not amenable to surgical repair. Other radiological studies have demonstrated that AHD is smaller in patients with RC tendinopathy¹⁴, is positively associated with the size of the RC tear and degree of fatty degeneration of the RC muscles¹⁵, and is a predictor of both short-term disability¹⁶ and functional status¹⁷.

Surgical interventions, such as acromioplasty, as well as many rehabilitation interventions for RC tendinopathy, are based around attempting to correct or ameliorate a reduced AHD, with the expectation that this will improve shoulder symptoms and function². It is therefore important that a reliable method of AHD measurement is identified, in order to confirm the veracity of this hypothesis. Reliability of a measurement relates to the degree to which it is consistent, and free from error. There are numerous variables that may influence the reliability of AHD measurement including; type of imaging used, measurement protocol, patient position, presence and degree of tendinopathy, and inter-examiner variables. Although a recent review evaluated AHD measurement by ultrasound in RC tendinopathy¹⁸, concluding that ultrasound-measured AHD is smaller in individuals with RC tears, no assessment of measurement reliability was carried out. No reviews to date have examined the reliability of any other radiological method. There is a need for a systematic review assessing the reliability of AHD measurement, so that a more robust basis for the assessment of AHD in individuals with RC tendinopathy can be recommended. In turn, the contribution of reduced AHD to shoulder pain and loss of function, and the impact of AHD alteration with physical or surgical interventions

could be determined. Therefore, the aim of this review was to systematically assess the evidence for the intra- and inter-rater reliability of radiological methods of measuring AHD in relation to RC tendinopathy.

METHODS

Inclusion/Exclusion criteria

Studies describing the reliability of a method of measuring AHD by any radiological method (specifically radiographs, MRI/MRA, CT, ultrasound) were the focus of this review. We included studies that reported collection and analysis of any reliability data, whether or not this was a primary aim. Studies involving human adult populations, either healthy subjects or subjects with diagnosed RC tendinopathy of any degree, as well as studies including those with RC tendinopathy as a subset of other shoulder pathologies were included. We excluded studies of patients with non-RC shoulder disorders e.g. instability, and neurological conditions, as the degree and direction of change in AHD is likely to be different in these populations. We did however conduct a sensitivity analysis to assess how many papers of non-RC disorders were excluded and whether this had any influence on the conclusions of this review. We only included studies published in the English language.

Search Strategy

The search strategy was developed with the help of a medical librarian, and involved searches of the following databases, from inception until June 2012: PubMed, CINAHL, MEDLINE, AMED, Sport Discus (using a combined search on the EBSCO database); Google Scholar; ProQuest digital dissertations; Cochrane Central Register of Controlled Trials (CENTRAL); and the Physiotherapy Evidence Database

(PEDro). Searches were conducted using the search terms and combinations illustrated in Fig 1.

The names of the radiological methods (e.g CT, MRI) were not used in the final search, as this may have restricted the number of papers identified. We also did not include terms related to reliability, as our aim was to also include papers where reliability analysis was conducted as a pilot or secondary aspect of the study. The search syntax was modified to match that in use in each of the databases.

The reference list of each relevant full-text article was reviewed to identify any potential additional references, as well as that of the single relevant systematic review identified. Initial screening of articles by title and abstract to remove clearly unrelated titles, was conducted by a single examiner. All identified references were examined by title and abstract in relation to the inclusion and exclusion criteria, independently by two reviewers (KM, JL). Potentially relevant articles were then obtained in full text format. Two reviewers applied the selection criteria to the full text articles, to determine the final ones to be included. A third examiner (JC) was available to resolve disagreement, however was not required. A citation search for all included studies was carried out, but no further relevant studies were identified.

Quality Assessment and Data extraction

Quality assessment of the included studies was completed using the Quality Appraisal for Reliability Studies (QAREL) checklist¹⁹. As recommended by Lucas et al¹⁹, piloting of the checklist was carried out as follows: a single study (one of those

included in the review) as jointly assessed by the examiners, with discussions and agreement as to how each item was to be defined in relation to this review. Then two studies excluded from this review (as they examined AHD measurement in neurological populations) were independently assessed with the checklist by two examiners (KM, JL). Agreement on the first study was 72%. Further discussions followed, and subsequently agreement on the second study was 100%.

QAREL checklist items are described in Table 1. It was deemed that item 5 of the checklist was not applicable to this review as there is currently no accepted, definitive reference standard for the measurement of AHD. In relation to the final item, regarding statistical measures, we required that studies using an intra-class correlation co-efficient (ICC) reported the model of ICC being used, and also that estimates of precision be presented, in order to achieve a “Yes score” for this item. Two reviewers independently assessed all included studies. Studies were deemed to be high quality if at least 50% of applicable items were rated as “Yes” on the checklist. Data was extracted from the studies using the QAREL extraction form. The appropriateness of the radiological protocols used in the included studies was separately assessed by an experienced musculoskeletal radiologist (JC). For the purposes of this review, the reliability estimates from ICCs were categorised as suggested by Fleiss²⁰ that is, >0.75 =excellent reliability, $0.40-0.75$ =fair to good reliability, and <0.40 =poor reliability. Due to the heterogeneous nature of the methods and populations studied, pooling of data was not deemed appropriate in this review. Instead a ‘levels of evidence’ approach was taken, using a modified version of the Cochrane Back Pain Group criteria²¹ i.e:

- Strong evidence –consistent findings in multiple high quality studies

- Moderate evidence – consistent findings in one high quality and one or more lower quality studies
- Limited evidence –consistent findings in one or more lower quality studies
- No evidence – if there were no studies or conflicting results

RESULTS:

The literature search retrieved a total of 2115 citations, from which 2073 non-relevant titles or duplicates were removed in the initial screening by a single examiner (See Figure 2). Two reviewers then assessed the remaining 42 by title and abstract. These were narrowed down to 21 based on the specified inclusion and exclusion criteria. Following the examination of full-text, a further two were obtained from the reference lists. Of these, 18 were included in the review, following exclusion of 5 papers as 3 contained reliability co-efficients only, without details of the method of reliability assessment, and 2 did not present the AHD measures separately, but as part of a ratio measure. In the sensitivity analysis, two papers evaluating AHD measurement reliability in non-RC shoulder disorders were identified. Both were carried out in populations with hemiplegic shoulders, and both concluded that ultrasound was a highly reliable method of measurement, aligning closely with the conclusions of the papers included in this review.

Quality assessment

Agreement for QAREL items between the two assessors after independent assessment of the 18 studies was 88%, with an average Cohen's kappa value of 0.95, demonstrating an excellent level of inter-rater agreement²². Joint discussions resolved remaining areas of disagreement, and the final ratings are displayed in Table 2. Of the 18 studies assessed with the QAREL checklist, 8 were deemed to be of high quality. There was limited detail describing the blinding of examiners, and randomisation of testing procedures, resulting in the majority of studies being rated "Unclear" on items 6, 7 and 8. A high proportion of incorrect statistical analysis, or inadequate information, also led to Item 11 being rated as "No" or "Unclear" for most studies. Six studies gave no information regarding training or experience of the testers, which left the rating for Item 2 "Unclear" for these studies. The decision to include studies which reported reliability data as a secondary aspect only, presented a possible risk of increasing the proportion of lower quality studies. However, there was in fact a similar proportion of high quality studies among both the primary (4/11) and secondary (3/7) reliability studies examined. There was also no difference in the proportion of high quality studies according to radiological modality.

Types of study

The details of the studies included in this review in terms of study type, population, testers, methods, reliability data, and mean AHD values reported, divided up according to radiological modality used are summarised in Tables 3-5 (supplementary file). Over half of the studies (10/18) employed US to assess AHD, with 4 studies using radiographs, 2 employing MRI only, and 2 using combined methods (MR and radiographs, or MRI, radiographs and CT). Eight studies assessed intra-rater

reliability - these were predominantly studies using US (7/8). Inter-rater reliability, or both types of reliability, were assessed in 5 studies each. The majority of studies (12/18) assessed reliability of AHD measurement in healthy or athletic populations, with 6 investigating people with shoulder pain.

Ultrasound studies

Of the 10 studies using ultrasound to measure AHD²³⁻³², 7 assessed the reliability of a single examiner (intra-rater), while one investigated reliability of 2 or more examiners (inter-rater), and 2 studied both types of reliability. Two studies included reliability data on participants with shoulder pain. Kalra et al²⁵ studied intra-rater reliability only in 31 participants with MRI-diagnosed RC disease (mean age: 53.5 years), and Pijls et al³⁰ studied both inter- and intra-rater reliability in 43 people described as having subacromial impingement syndrome (mean age: 51 years) as diagnosed by an orthopaedic surgeon, without giving any details of how diagnosis was determined. The remaining 8 studies assessed reliability in pain-free participants. Mean subject age in the pain-free study groups ranged between 21 and 34 years, with a single study by Kumar et al²⁶ involving older participants (mean age 64.2 years). Five studies were deemed to be of high methodological quality (Table 2), and each of these reported a good or excellent level of either inter and/or intra-rater reliability.

The testers undertaking the ultrasound scanning were physiotherapists in 7 of the studies^{24,26-9,31-32}, variously described as having training ranging from one hour to 3 months; and radiologists in two of the studies^{23,30}, with one study not providing details of the raters²⁵. One study reported similar degrees of reliability between an

experienced radiologist (ICC=0.94) and a novice in ultrasound (ICC=0.92)³⁰, while in two studies by the same authors, one using a physiotherapist trained in shoulder ultrasound, and the other using student physiotherapists, with limited training or experience, slightly better reliability was reported with the experienced examiner (ICC=0.96-0.99)²⁶ compared to the novices (ICC=0.88-0.91)²⁷.

All studies reported the reliability of measuring separate images of the same subject, while two also studied the intra-rater reliability of repeated measurements of the same image^{30,32}. There were varied time intervals used for the repeated scans, ranging from within-session scans, to others taken up to two weeks²⁶, or 6 weeks later²⁹. While various factors may influence the normal variation in AHD over time (such as posture, fatigue, activity), no studies described controlling for these factors.

All studies used a high frequency linear transducer (between 5 and 12.5 MHz) to acquire the ultrasound scans. There was variation in transducer placement, with two studies placing it on the anterior part of the acromion^{23,31}, while others used the posterior or mid-acromion²⁵, or did not give adequate details of the testing protocol. There were also differences in how AHD was measured between studies. Six studies described that the measurement of the shortest distance between the acromion and humeral head was assessed, usually along a line parallel to the acoustic shadow cast by the acromion^{23, 25,29-32}. In contrast, 3 studies measured the distance between edge of the acromion and the tip of the greater tuberosity²⁶⁻⁸, which anatomically is a longer distance. Duerr²⁴ reported equal reliability measuring both of these distances. As the greater tuberosity cannot be visualised when the arm is moved into abduction, the author recommended the alternative measurement as the standard. The neutral

shoulder position was used in all studies, while additional scans in various positions of either active or passive abduction (30°, 45°, 60° and 90°) were carried out by 7 studies (Table 3 - supplementary file).

Overall there was a strong level of evidence for the reliability of ultrasound in the measurement of AHD. Intra-rater reliability was found to be good to excellent with almost all ICC values being above 0.75. However inter-rater reliability was poorer, with the single high quality inter-rater study³⁰ reporting an inter-rater ICC of just 0.70. Since study methods were similar across the ultrasound studies, forest plots were constructed to illustrate the range of ICC values reported (Figures 3 and 4). Standard error of measurement values for AHD were more variable; they were below 1mm in the high quality studies by Dueri²⁴ and Seitz³¹, and in a number of lower quality studies^{26,28-29}, however Kalra et al²⁵ reported higher SEM values of 0.9-1.6mm in their high quality study. Three studies reported the minimal detectable change (MDC) for AHD measurement, which is an important concept representing the amount of change required to exceed measurement variability. Reported MDC values in neutral shoulder position for AHD were 0.9mm²⁴ and 1.3mm²⁵, and 2.1mm for acromion to greater tuberosity distance²⁸.

Radiograph studies

Among the 6 studies assessing the reliability of using radiographs to measure AHD, 4 included people with shoulder pain^{15,33,35,37} (with two including people with confirmed rotator cuff disease^{15,37}), and a further two solely using pain-free

participants^{34,36}. The mean age range of the shoulder pain subjects was from 55-59 years, while the pain-free groups were in the 20-35 year age range (See Table 4-supplementary file). Two studies each assessed either intra-rater^{34,36} or inter-rater reliability^{15,37}, while the remaining two studied both types of reliability^{33,35}. Three of the radiograph studies were deemed to be of high methodological quality^{15,34,35}.

Three studies examined the reliability of radiographs in measuring AHD, one examined digital fluoroscopy³⁶, and two studied radiographs along with other modalities^{15,37}. Three of the radiograph studies used standardised views^{15,35,37}, while Bernhardt et al³³ examined non-standardised films from various clinics, and Fehringer et al³⁴ studied the effect of differing beam angles. The majority of the studies examined the reliability of reading a single set of radiographs, whereas the Thompson et al³⁶ and Fehringer et al³⁴ studies measured the reliability of reading repeated films of the same participant; with each of the radiographs taken at a different angle or arm position Fehringer et al³⁴ study. Poor reliability was reported for the measurement of these different views. In the studies by Bernhardt et al³³ and Gruber et al³⁵, it was clear that reliability was enhanced when using standardised radiographs (intra-rater: maximum difference = 3mm), prospectively collected for the study, than when non-standardised radiographs were examined retrospectively (intra-rater: maximum difference = 7mm). In one of the high quality studies, Saupe et al¹⁵ examined the inter-rater reliability of measuring AHD on standard AP radiographs between an experienced and non-experienced radiologist, and reported excellent reliability (ICC= 0.77), similar to the findings of the poorer quality Werner et al³⁷ study, where four observers also achieved excellent inter-tester reliability examining AHD on 40 radiographs. Thompson et al³⁶ reported good intra-rater reliability of measuring digital

fluoroscopy images taken immediately in succession (ICC=0.75-0.99), with poorer reliability for those taken 9 months apart (ICC=0.3-0.99). It was not possible to undertake any direct comparisons of the reliability data between these studies, as a wide variety of statistical methods of reporting were used. However, overall, due to the use of non-standardised imaging, or poor reporting of statistical analysis, the evidence was conflicting for the reliability of AHD measurement using radiographs, which according to the Cochrane criteria, equates to no evidence.

CT and MRI studies

Two studies examined the measurement of AHD using open MRI systems^{38,39} while one used conventional MRI¹⁵, and another both MRI and CT imaging³⁷. A single study, by Saupe et al¹⁵, was rated as being of high quality.

The method of measurement was similar across all studies using CT or MR imaging, using the shortest distance between the inferior surface of the acromion and the upper sub-chondral surface of the humeral head (See table 5 – supplementary file). Two studies examined inter-rater reliability^{15,37}. Two also examined intra-rater reliability, with Hinterwimmer et al³⁸ achieving this by repeating the scans within a single session, while Moffet et al³⁹ did so by re-reading the same scans one month later.

For the open MRI studies, only pain-free subjects were used. Hinterwimmer et al³⁸ used a single healthy volunteer for inter-rater reliability assessment and reported low co-efficients of variation, suggesting reasonable accuracy of measurement, while Moffet et al³⁹ assessed both inter and intra-rater reliability in 13 pain-free subjects,

and reported excellent reliability, with ICCs all >0.75. People with shoulder pain participated in the studies by Werner et al³⁷ assessing the reliability of AHD measures with MRI, CT, and radiographs, and Saupe et al¹⁵ using MRI and radiographs. MRI and CT imaging were shown to be similarly reliable to radiographs in the Werner et al³⁷ study, while MRI had better reliability than radiographs in the Saupe et al¹⁵ study. Since Werner et al³⁷ reported their reliability statistics using regression analysis ($r=0.8$ for CT and MRI), and Saupe et al¹⁵ reported an ICC value (0.91), it is difficult to make direct comparisons of the degree of reliability, however it appears that good levels of inter-rater reliability was achieved in both studies. Overall, there was a moderate level of evidence for the reliability of AHD measurement using CT and MR imaging (based on results of one high and two lower quality studies).

DISCUSSION

This review evaluated the reliability of AHD measurement using radiological means. The majority of reliability studies assessed ultrasound methods. Study quality, as assessed by the QAREL checklist, was generally poor with less than half of the studies deemed to be of high quality. Major weaknesses of the studies reviewed were in the areas of tester blinding to their own and others measures, as well as to additional cues, such as side of symptoms, while undertaking imaging and/or measurement. While overall levels of reliability were good to excellent across the studies, there was more high quality evidence for the reliability of ultrasound as a method of AHD measurement, than for other modalities.

When assessing the reliability of imaging-based assessments, two distinct aspects of reliability exist. One is the reliability of measuring the image itself, incorporating any variability associated with localising anatomical landmarks, how measurements are made, and measurement error; and which is assessed by carrying out repeated measurements of the same image. The other is the reliability of taking repeated images of the same subject, which encompasses a myriad of variables such as; positioning of the subject, setting imaging parameters, operator-related variability, machine calibration, etc. The first type is likely to yield better reliability co-efficients, with less potential for variation. The second type is more challenging, potentially yielding poorer reliability; however it is important in the context of test-retest study. All of the ultrasound studies assessing intra-rater reliability assessed the measurement of repeated images, while some also undertook re-measurement of the same images. However the majority of the other imaging studies re-measured the same set of images, which may have led to over-inflation of reliability levels. Thompson et al³⁶ and Hinterwimmer et al³⁸ carried out repeated imaging (in digital fluoroscopy and open MRI respectively), however this was in a very small number (N=1 and N=5) of pain-free subjects. It is accepted however, that repeated radiation exposure may make this type of study ethically unacceptable for radiograph and CT studies.

Item 9 in the QAREL checklist (Table 1) emphasises the importance of taking into account the stability of the measure when determining the time scale for repeated measures¹⁹. The allowance of a significant time lapse between testing sessions, without indicating how possible confounding variables such as fatigue and posture, have been controlled for, threatens the internal validity of the observations being made. In contrast, the studies in this review using time intervals of a few days up to 6

weeks in pain-free subjects reported excellent reproducibility of the AHD measures, despite little description of controlling for confounding variable, suggesting that the measurement is reasonable stable over this time period. However, only within-session reliability of AHD measurement was available in this review for shoulder pain populations, therefore the stability of the measure in RC pathology over time is unknown.

The widespread use of pain-free populations for reliability testing reduces the external validity of the findings, as there may be significant differences in the degree of reliability achieved in pain-free versus shoulder pain populations due to greater challenges of positioning the painful arm and the potential influence of pathology on image quality. One third of the included studies involved people with shoulder pain, and one study²⁹ included an athletic population. Within the pain-free populations, the age range of subjects tended to be much lower than the typical population age range for RC disorders. The studies by Kumar et al^{26,27} demonstrated the importance of including the relevant age groups as controls, due to the lower AHD values in the older age group. In the studies of shoulder pain groups, Kalra et al²⁵ and Saupe et al¹⁵ confirmed the RC pathology using imaging, with Pijls et al³⁰ using a clinical diagnosis, and Werner et al³⁷ using unspecified diagnostic criteria, and a non-specific shoulder pain group used in the remaining studies^{33,35}. Unfortunately, the single study to include both shoulder pain subjects and pain-free controls did not separately report the reliability for the two groups²⁵. Further information is also required concerning AHD reliability in athletic versus non-athletic populations.

The issue of tester experience and qualifications is important in any imaging-based study. Ultrasound, in particular, is said to be highly operator-dependant⁴⁰. Two of the ultrasound studies stated that radiologists conducted the scans, while the remaining majority stated that operators were physiotherapists with varying, but generally poorly described levels of training and experience in shoulder ultrasound imaging. An experienced and novice ultrasound operator achieved similarly excellent reliability in the Pijls et al³⁰ study, while the two studies by Kumar^{26,27} illustrate better reliability for an operator with moderate levels of training versus physiotherapy students (although both studies achieved ICC >0.75). However, these studies suggest that reliable AHD measures can be achieved with a limited level of ultrasound training, in contrast to the higher level of training and experience needed to accurately undertake full diagnostic assessment of the shoulder. Based on MDC values provided, a single tester can achieve an accuracy level of between 0.9-1.3mm for ultrasound measurement of AHD in neutral, so that any change beyond this can be accepted as true change. No MDC was reported in the inter-tester reliability studies.

In addition to tester variability, the standardisation of imaging protocols for radiographs, CT and MRI are important elements in assessing reliability, clearly evidenced in the Bernhardt et al³³ and Fehringer et al³⁴ studies, where the use of non-standardised radiographs was shown to negatively influence reliability. As previously discussed in the review by Seitz et al¹⁸, different landmarks were used for the measurement of AHD in some of the ultrasound reliability studies. Most authors measured the shortest distance between the inferolateral acromion and the closest part of the humeral head, however others chose to measure AHD from the acromion to the tip of the greater tuberosity²⁶⁻²⁸, while Duerr²⁴ carried out both measures. The shortest

distance measurement is most closely aligned with the measurement protocols used in radiograph, MR and CT imaging studies, therefore it is suggested that this is the most useful measurement to report. Overall, it is recommended that clearer descriptions of tester experience, and standardisation of imaging procedures are provided in imaging reliability studies, to allow better extrapolation of findings across studies.

While reliability of a measurement is critical to inform its use, since it refers to how consistent a measuring device is, it is equally important to evaluate the validity of the measurement, which confirms whether a study measures or examines what it claims to measure. As the sub-acromial space is a 3-dimensional space, there is an inherent problem in that, conventional radiographic imaging merely measures in two dimensions. There is no evidence or general agreement as to which modality provides a “gold standard” for AHD measurement, which led to the exclusion of Item 5 of the QAREL checklist for this review. With ultrasound, it is not possible to view the under-surface of the acromion due to the acoustic shadow produced by the bone; therefore the area of the smallest AHD may not be viewed or measured accurately. While radiographs provide a clearer view of the bony structures, projection issues and bony overlap may lead to measurement inaccuracies. Both ultrasound and radiographs are conducted in the functional upright position, adding face validity to these measures, however standard MR and CT imaging is carried out with the patient in the supine position, where the lack of the arm weight, gravity and absence of muscle activity may lead to lower AHD values being measured. This was observed in the studies by Werner et al³⁷ and Saupe et al¹⁵, where inter-method comparisons were carried out between radiographs and CT or MRI. Saupe et al¹⁵ reported a poor correlation between AHD measured on radiographs and MR images, with AHD

values of 2.8mm less on average for the MR images. Similar differences between radiographs and MRI were noted by Werner et al³⁷, who used linear regression to provide a conversion formula. While a full discussion of the issues relating to validity of AHD measurement is beyond the scope of this review, it is important that validity is given further consideration before AHD measures are more widely used in diagnosis or treatment.

In summary, ultrasound measurement of AHD demonstrated sufficient intra-rater reliability in healthy populations and in two cohorts of patients with RC tendinopathy to be the recommended method of AHD measurement. However, because low ICC values were reported in the single high quality inter-tester study³⁰, an additional study of people with imaging-confirmed RC pathology is required to ascertain the inter-tester reliability in this population. The evidence for reliability of AHD measured by CT and MRI was moderate, with a number of studies demonstrating good to excellent reliability, generally derived from re-reading of a single set of images. The evidence for radiographs was conflicting, with the use of non-standardised images making comparisons difficult. As the cheapest and most accessible method, with no radiation exposure concerns, and where excellent reliability can be achieved with limited training, ultrasound is the recommended method of AHD measurement.

Limitations

This review was comprehensive, including a variety of published sources e.g. peer-reviewed publications, and theses, however we did not search extensively for grey literature, which may have limited the number of studies. We only included English-language papers, however no relevant papers were excluded due to the language restriction. While a lack of information in the published papers caused some studies to be rated as “Unclear” on a number of QAREL items, we did not contact authors for

further information, as it was deemed that such a process may be subject to excessive recall bias. The use of the QAREL checklist in this review provided a standardised method of quality assessment¹⁹. The piloting process was important in the resulting high level of agreement reached between reviewers in this review. However, the QAREL checklist is a relatively new quality assessment tool, and as yet no published studies have reported on its reliability, or validity for use in systematic reviews of diagnostic tests. Further testing of its psychometric properties is required before it can be broadly recommended for use. The extent of conclusions that could be reached was limited by the generally low quality of the included studies.

Conclusion

This review found that intra-rater reliability of AHD measurement by ultrasound is well supported in healthy populations, while also highlighting the scarcity of high quality studies in people with RC pathology, and inter-tester reliability studies. There was moderate evidence for the reliability of AHD measurement with CT and MRI, and no evidence for the reliability of radiographic methods. Based on the evidence reviewed, ultrasound is the authors' recommended method of AHD measurement, however further data on inter-rater reliability in symptomatic populations is required. With regard to MR and CT; improved standardisation of methods, and assessing the reliability of both repeated imaging and image re-measurement should be considered, to provide a solid basis for the using these methods to measure AHD. At present, radiographs are not recommended for AHD measurement as there is no evidence to support their reliability.

What this study adds:

- This is the first systematic review of the reliability of radiological methods of acromio-humeral distance measurement
- The measurement of acromio-humeral distance using ultrasound is highly reliable for a single tester, while inter-rater reliability requires further investigation
- While there is moderate evidence to support the reliability of acromio-humeral distance measurement using CT and MRI, the reliability of radiographic methods has not been substantiated

Acknowledgements: none

Competing Interests: none

Funding: The first author (KM) is funded under a Research Fellowship from the Health Research Board of Ireland.

Contribution:

KM and JL were involved in conception and design. KM and JL independently reviewed the literature. KM and JL extracted the study data. KM and JL were involved in rating the literature, with JC acting to mediate disagreements in ratings. All authors were involved in data analysis and interpretation, as well as preparing the manuscript for publication

Reference list

1. Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004; **33**(2):73-81.
2. Lewis JS Rotator cuff tendinopathy: a model for the continuum of pathology and related management. *Br J Sports Med* 2010; **44**:918–923
3. Deutsch A, Altchek DW, Schwartz E, et al. Radiologic measurement of superior displacement of the humeral head in the impingement syndrome. *J Shoulder Elbow Surg.* 1996; **5**(3):186–193
4. Neer CS. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *J Bone Joint Surg [Am]* 1972; **54**:41–50.
5. Ludewig PM, Reynolds JF. The association of scapular kinematics and glenohumeral joint pathologies. *JOSPT* 2009; **39**:90-104
6. Ainsworth R, Lewis JS, Conboy V A prospective randomised placebo controlled clinical trial of a rehabilitation programme for patients with a diagnosis of massive rotator cuff tears of the shoulder. *J Shoulder Elbow Surg* 2009;**1**:55-60

7. Schmidt, W. A., Schmidt, H., Schicke, B, et al. Standard reference values for musculoskeletal ultrasonography. *Ann Rheum Dis* 2004; **63**:988-994
8. Wang HK, Lin JJ, Pan SL, et al. Sonographic evaluations in elite college baseball athletes. *Scand J Med Sci Sports*. 2005; **15**:29-35
9. Girometti R, De Candia A, Sbuelz M, et al. Supraspinatus tendon US morphology in basketball players: correlation with main pathologic models of secondary impingement syndrome in young overhead athletes. *Radiol Med*. 2006;**111**:42-52.
10. Giphart JE, van der Meijden OA, Millett PJ. The effects of arm elevation on the 3-dimensional acromiohumeral distance: a biplane fluoroscopy study with normative data. *J Shoulder Elbow Surg*. 2012; **21**:1593-1600
11. Graichen H, Bonel H, Stammberger T, et al. Sex-specific differences of subacromial space width during abduction, with and without muscular activity, and correlation with anthropometric variables. *J Shoulder Elbow Surg* 2001;**10**:129–135
12. Chopp JN, O'Neill JM., Hurley K, et al. Superior humeral head migration occurs after a protocol designed to fatigue the rotator cuff: A radiographic analysis. *J Shoulder Elbow Surg* 2010;**19**:1137-1144
13. Goutallier D., Le Guilloux P, Postel JM, et al. Acromio- humeral distance less than six millimeter: Its meaning in full-thickness rotator cuff tear. *Orthop Traumatol Surg Res* 2011;**97**:246-251

14. Cholewinski JJ, Kusz DJ, Wojciechowski P, et al. Ultrasound measurement of rotator cuff thickness and acromio-humeral distance in the diagnosis of subacromial impingement syndrome of the shoulder. *Knee Surg Sports Traumatol Arthrosc* 2008;**16**:408–14
15. Saupe N, Pfirrmann CWA, Schmid MR, et al. Association between rotator cuff abnormalities and reduced acromiohumeral distance. *Am J Roentgenol* 2006;**187**:376–38
16. Hebert LJ, Moffet H, Dufour M, et al. Acromiohumeral distance in a seated position in persons with impingement syndrome. *J Magn Reson Imaging* 2003;**18**:72–79
17. Mayerhoefer ME, Breitenseher MJ, Wurnig C, et al. Shoulder impingement: relationship of clinical symptoms and imaging criteria. *Clin J Sport Med* 2009;**19**:83–89
18. Seitz AL, Michener LA. Ultrasonographic measures of subacromial space in patients with rotator cuff disease: a systematic review. *J Clin Ultrasound* 2011;**39**:146–54
19. Lucas NP, Macaskill P, Irwig L, et al. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL) *J Clin Epidemiol* 2010;**63**:854–861

20. Fleiss JL. *The Design and Analysis of Clinical Experiments*. 1986, New York: Wiley
21. van Tulder M, Furlan A, Bombardier C, et al. Updated method guidelines for systematic reviews in the Cochrane Collaboration back review group. *Spine* 2003;**28**:1290–9.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159-74
23. Desmeules F, Minville L, Riederer B, et al. Acromio-humeral distance variation measured by ultrasonography and its association with the outcome of rehabilitation for shoulder impingement syndrome. *Clin J Sport Med* 2004;**14**:197-205
24. Duerr M. Reliability and accuracy of distance measurements between shoulder bony landmarks evaluated by ultrasound in asymptomatic subjects. 2010, Masters Thesis, Auckland University of Technology, New Zealand
Accessed June 20th, 2012 on <http://aut.researchgateway.ac.nz/handle/10292/1109>
25. Kalra N, Seitz AL, Boardman ND, et al. Effect of posture on acromiohumeral distance with arm elevation in subjects with and without rotator cuff disease using ultrasonography. *JOSPT*, 2010;**40**:633-40

26. Kumar P, Bradley M, Swinklels A. Within-day and day-to-day intrarater reliability of ultrasonographic measurements of acromion-greater tuberosity distance in healthy people. *Physiother Theory Pract*. 2010;**26**:347-51.
27. Kumar P, Chetwynd J, Evans A. Interrater and intrarater reliability of ultrasonographic measurements of acromion-greater tuberosity distance in healthy people. *Physiother Theory Pract* 2011;**27**:172-5.
28. Leong HT, Tsui S, Ying M, et al. Ultrasound measurements on acromio-humeral distance and its correlations with supraspinatus tendon thickness and shoulder rotational strengths. *J Sci Med Sport*. 2012;**15**:284-91
29. Maenhout A, van Cingel R, de Mey K, et al. Sonographic evaluation of the acromiohumeral distance in elite and recreational female overhead athletes. *Clin J Sport Med* 2012 Advance online publication, doi: 10.1097/JSM.0b013e31825b6995
30. Pijls BG, Kok FP, Penning LIF, et al. Reliability study of the sonographic measurement of the acromiohumeral distance in symptomatic patients. *J Clin Ultrasound*, 2010;**38**:128-134
31. Seitz AL, McClure PW, Finucane S, et al. The scapular assistance test results in changes in scapular position and subacromial space but not rotator cuff strength in subacromial impingement. *JOSPT*, 2012;**42**:400-12.
32. White CE, Dedrick GS, Apte GG, et al. The effect of isometric shoulder internal and external rotation on the acromiohumeral distance. *Am J Phys Med Rehabil*. 2012;**91**:193-9.

33. Bernhardt GA, Glehr M, Zacherl M, et al. Observer variability in the assessment of the acromiohumeral interval using anteroposterior shoulder radiographs. *Eur J Orthop Surg Traumatol*, 2012, Advanced online publication, DOI: 10.1007/s00590-012-0942-y
34. Fehringer EV, Rosipal CE, Rhodes DA, et al. The radiographic acromiohumeral interval is affected by arm and radiographic beam position. *Skeletal Radiol*. 2008;**37**:535-9.
35. Gruber G, Bernhardt GA, Clar H, et al. Measurement of the acromiohumeral interval on standardized anteroposterior radiographs: a prospective study of observer variability. *J Shoulder Elbow Surg* 2010;**19**:10-3.
36. Thompson MD, Landin D, Page PA. Dynamic acromiohumeral interval changes in baseball players during scaption exercises. *J Shoulder Elbow Surg* 2011;**20**: 251-8.
37. Werner CM, Conrad SJ, Meyer DC, et al. Intermethod agreement and interobserver correlation of radiologic acromiohumeral distance measurements. *J Shoulder Elbow Surg* 2008;**17**:237-40.
38. Hinterwimmer S, Von Eisenhart-Rothe R, Siebert M, et al. Influence of adducting and abducting muscle forces on the subacromial space width. *Med Sci Sports Exerc* 2003;**35**:2055-9.

39: Moffet H, Hebert LJ, Dufour M, et al. Variation in subacromial distance measured by MRI during shoulder flexion and abduction movements. *Can J Rehabil* 1998;**11**:265-267.

40. O'Connor PJ, Rankine J, Gibbon WW, et al. Interobserver variation in sonography of the painful shoulder. *J Clin Ultrasound* 2005;**33**: 53–56

Table 1: QAREL checklist items

QAREL CHECKLIST
<ol style="list-style-type: none"> 1. Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? 2. Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? 3. Were raters blinded to the findings of other raters during the study? (<i>inter-rater studies only</i>) 4. Were raters blinded to <i>their own</i> prior findings of the test under evaluation? (<i>Intra-rater studies only</i>) 5. Were raters blinded to the subjects' disease status or the results of the accepted reference standard for the target disorder (or variable) being evaluated? (<i>Excluded in this review</i>) 6. Were raters blinded to clinical information that was not intended to form part of the study design or testing procedure? 7. Were raters blinded to additional cues that are not part of the test? 8. Was the order of examination varied? 9. Was the stability (or theoretical stability) of the variable being measured taken into account when determining the suitability of the time interval among repeated measures? 10. Was the test applied correctly and interpreted appropriately? 11. Were appropriate statistical measures of agreement used?

Table 2: Quality appraisal of included studies using QAREL checklist

Study	Method	Study type	QAREL ITEMS											*High Quality?
			1	2	3	4	6	7	8	9	10	11		
Bernhardt et al 2012	XRay	1°	?	?	?	?	?	?	?	N	Y	?	No	
Desmeules et al 2004	US	1°	N	Y	?	NA	?	?	Y	Y	Y	?	No	
Duerr 2010	US	1°	Y	?	NA	Y	NA	Y	?	Y	Y	Y	Yes	
Fehringer et al 2008	XRay	1°	Y	Y	Y	NA	Y	Y	Y	Y	Y	?	Yes	
Gruber et al 2010	XRay	1°	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Yes	
Hinterwimmer et al 2003	MRI	2°	Y	?	?	?	?	?	?	?	Y	N	No	
Kalra et al 2010	US	2°	Y	?	NA	Y	?	?	?	Y	Y	Y	Yes	
Kumar et al 2010	US	1°	Y	?	NA	?	N	?	?	?	Y	?	No	
Kumar et al 2011	US	1°	N	?	Y	Y	?	?	N	Y	Y	Y	Yes	
Leong et al 2012	US	1°	?	Y	NA	?	?	?	?	N	Y	N	No	
Maenhout et al 2012	US	2°	?	Y	NA	?	?	?	Y	?	Y	?	No	
Moffet et al 1998	MRI	1°	Y	Y	Y	?	NA	?	?	Y	?	?	No	
Pijls et al 2010	US	1°	?	Y	Y	Y	?	?	?	Y	Y	?	Yes	
Saupe et al 2006	XRay MRI	2°	?	Y	Y	NA	?	?	Y	Y	Y	?	Yes	
Seitz et al 2012	US	2°	Y	Y	NA	?	?	N	Y	Y	Y	Y	Yes	
Thompson et al 2011	XRay	2°	?	?	?	?	?	?	?	N	Y	?	No	
Werner et al 2008	CT MRI	1°	?	?	Y	?	?	?	?	Y	?	?	No	
White et al 2012	US	2°	Y	?	NA	?	?	?	Y	?	Y	?	No	

QAREL, Quality Appraisal for Reliability Studies; US, ultrasound

* For high quality, 50% of applicable items had to achieve a "yes" score

Y=Yes, N=No, NA=Not applicable, ?=Unclear

QAREL Item 5 was omitted, as it was considered not applicable due to there being no accepted reference standard for AHD measurement.

Study type: 1°= Reliability assessment was a primary aim of the study. 2°= study reported reliability data as a secondary or pilot aspect.

FIGURE LEGENDS

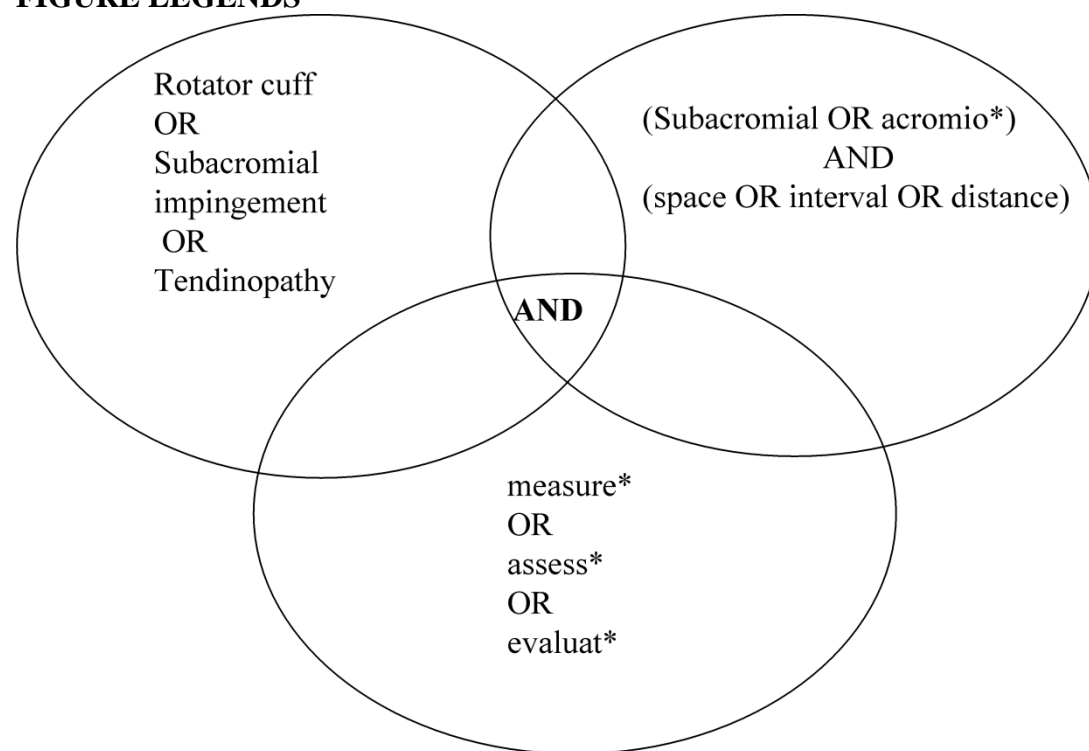


Figure 1: Boolean logic of search terms

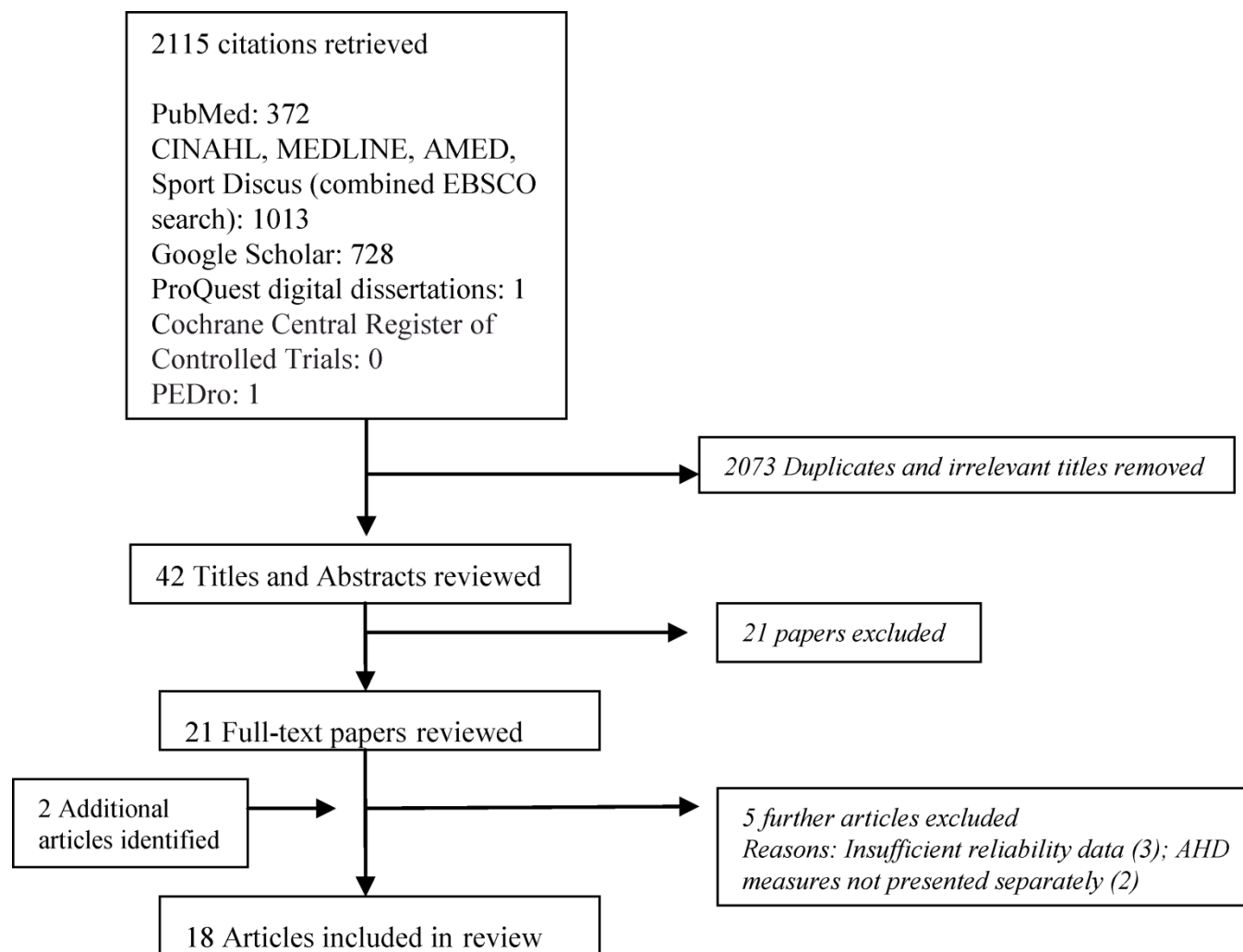


Figure 2: PRISMA Flow diagram of search results

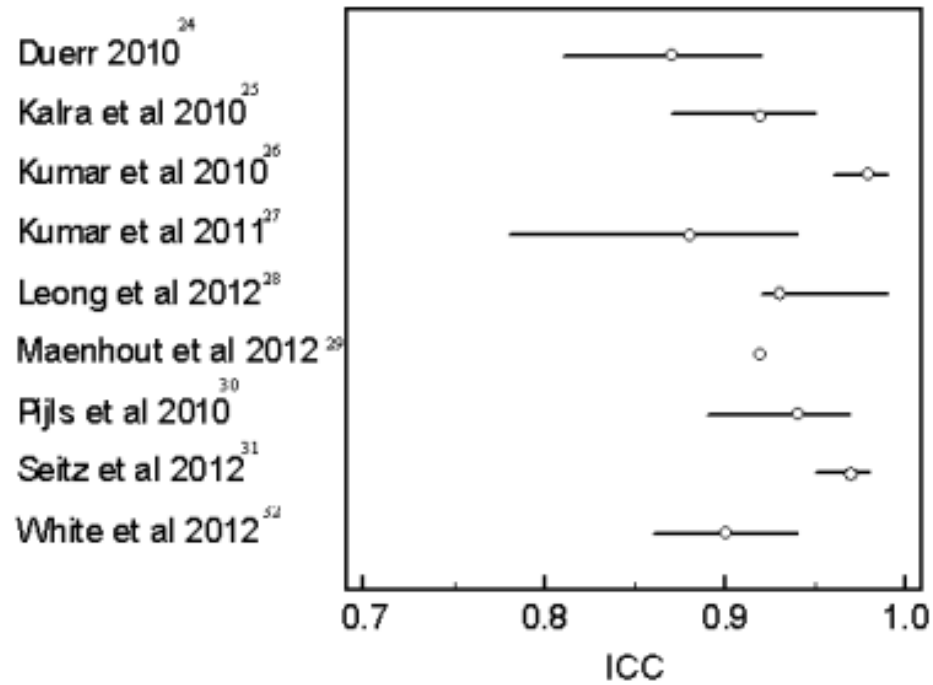


Figure 3: Illustrative forest plot of intra-rater reliability of acromiohumeral distance measurement with ultrasound in neutral shoulder position (ICCs +/- 95% confidence intervals, where available)

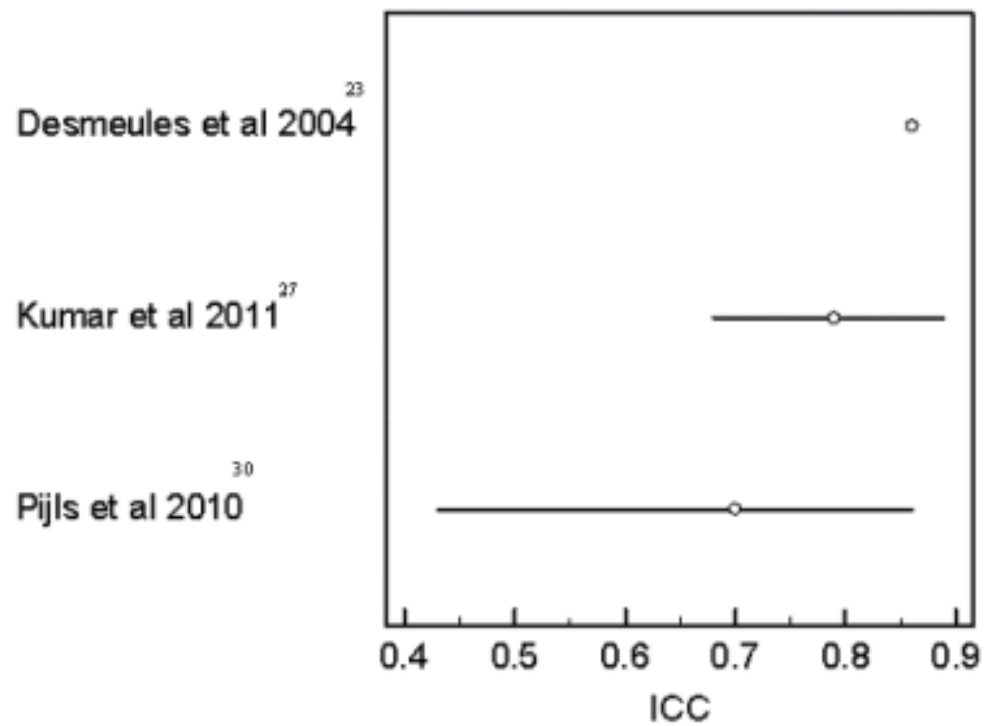


Figure 4: Illustrative forest plot of inter-rater reliability of acromiohumeral distance measurement with ultrasound in neutral shoulder position (ICCs +/- 95% confidence intervals, where available)

